Basis set dependence of the doubly hybrid XYG3 functional

Igor Ying Zhang,[1,2] Yi Luo,[2] and Xin Xu[1,3,a)]
[1]*State Key Laboratory of Physical Chemistry of Solid Surfaces, College of Chemistry and Chemical Engineering, Xiamen University, Xiamen 361005, China*
[2]*Department of Theoretical Chemistry, School of Biotechnology, Royal Institute of Technology, S-10691 Stockholm, Sweden*
[3]*Department of Chemistry, MOE Laboratory for Computational Physical Science, Fudan University, Shanghai 200433, China*

(Received 26 May 2010; accepted 23 August 2010; published online 10 September 2010)

In the present study, we investigated the basis set dependence of XYG3, a newly developed doubly hybrid functional [Y. Zhang, X. Xu, and W. A. Goddard III, Proc. Natl. Acad. Sci. U.S.A. **106**, 4963 (2009)], in prediction of (1) heats of formation (HOFs), (2) bond dissociation enthalpies (BDEs), (3) reaction barrier heights (RBHs), and (4) nonbonded interactions (NBIs). We used basis sets of triple-zeta quality starting from 6-311+G(d,p) with increasing completeness of the polarization functions to the largest Pople-type basis set 6-311++G(3df,3pd) and found that there was a continued improvement with larger basis sets. We showed that while HOF predictions were prone to basis set deficiencies, the basis set dependences in calculating BDEs, RBHs, and NBIs were mild. All of them converged fast with the increase of basis set size. We extended XYG3 to propose the XYG3o functional which was specifically optimized for a particular basis set in order to enhance its performance when using basis set of moderate size. With the 6-311+G(2df,p) basis set, XYG3o led to MADs of 2.56 kcal/mol for HOFs of the G3/99 set, 1.17 kcal/mol for BDEs of the BDE92/07 set, 1.11 kcal/mol for RBHs of the NHTBH38/04 and HTBH38/04 sets, and 0.40 kcal/mol for NBIs of the NCIE31/05 set, being comparable to those obtained by using XYG3/6-311++G(3df,3pd). © *2010 American Institute of Physics*. [doi:10.1063/1.3488649]

I. INTRODUCTION

Density functional theory (DFT) (Ref. 1) has become the leading tool for molecular quantum chemical calculations in recent years due to its low computational cost in conjunction with good precision in calculating various physical and chemical properties of molecules. In practical applications of DFT in the frame work of Kohn–Sham (KS) method,[2] two pre-decisions have to be made. First, an approximate exchange-correlation functional has to be chosen, which is designed to approximate the exact, yet unknown, density functional. Second, a finite set of basis functions has to be chosen, which is used to expand the KS orbitals to approximate the exact ground state density and other related quantities.

The success of the KS implementation of DFT critically depends on the quality of the exchange-correlation functional. With different philosophies in their approximations, various DFT functionals have been developed.[2–29] A foundation of most approaches is the local density approximation (LDA) based on solutions of the uniform electron gas.[2–4] This is the first rung of the hierarchy of various DFT approximations,[29] which yields results of good or moderate accuracy for properties such as lattice constants, bulk moduli, equilibrium geometries, and vibrational frequencies, whereas LDA leads to bond energies and cohesive energies that are far too large.[30] The second and third rungs are the generalized gradient approximation (GGA) and meta-GGA, respectively.[29] GGAs[5–8] and meta-GGAs[9,10] significantly reduce the overbinding tendency of LDA, but generally remain inadequate for thermochemistry of molecules.[30] The development of the fourth rung functional (i.e., hybrid methods) is a big step toward greater accuracy by introducing some amount of "exact exchange" on the basis of the adiabatic connection formula.[11,19,31] The most popular hybrid GGA is B3LYP,[3,4,6,11,32] which is unfortunately poor for the predictions of noncovalent bonding interactions[21] and reaction barrier heights,[21] with performance degrading dramatically as system sizes increase.[33,34] A recent important development of hybrid meta-GGAs is the proposal of the M06 family of functionals,[22] which currently provide the highest accuracy with a broad applicability for chemistry. The final fifth rung of Jacob's ladder utilizes the unoccupied KS orbitals in addition to the occupied KS orbitals (e.g., doubly hybrids).[23–28] This final rung is expected to allow the heaven of chemical accuracy to be achieved for broad applications.[29] One has to note that the main problem with DFT methods is the lack of a well-defined method for systematically improving the exchange-correlation functional toward the exact limit. As many of the approximate functionals include several empirical parameters that are chosen based on fitting to experimental data, a higher rung functional does not necessarily guarantee a better performance than a lower rung functional for a specific property. The choice of a proper functional demands careful validations and good experiences.

Basis set is the other user defined component of a DFT

a)Author to whom correspondence should be addressed. Electronic addresses: xinxu@xmu.edu.cn and xxchem@fudan.edu.cn.

calculation.[27,34–38] Similar to that of the Hartree–Fock (HF) method, it is generally agreed that the basis set used in DFT is relatively easy to saturate.[36] Fairly accurate results may be obtained even with basis sets as small as double-zeta (DZ) plus polarization,[36] and a basis set of triple-zeta (TZ) spdf +diffuse quality is already quite closed to the KS basis set limit for conventional DFT calculations.[36] Hence for application purposes a polarized DZ type basis set is typically used,[39] and for functional development purposes a polarized TZ type basis set is often employed.[12–14,20] Dynamical correlation energy in terms of molecular orbital (MO) theory is well-known slowly convergent.[40] Thus, a doubly hybrid functional that includes a portion of MP2-like correlation is expected to suffer from a similar slow basis set convergence as in MO theory. In fact, in their development[25] of B2T-PLYP (T for thermodynamics), B2K-PLYP (K for kinetics), and B2GP-PLYP (GP for general purpose), an aug-pc3 (Ref. 36) basis set was employed, which is of quadruple-zeta (QZ) to quintuple-zeta (5Z) spdfg+diffuse quality. In application of Grimme's original B2PLYP,[24] a very large CQZV3P basis set [QZ with three sets of valence polarization functions (3P) plus core-polarization functions] was recommended. With CQZV3P, a mean absolute deviation (MAD) of 2.5 kcal/mol was claimed[24] for heats of formation (HOFs) of the G3/99 set,[41] while with standard Pople's 6-311+G(3df,2p),[42] we found B2PLYP degraded, leading to MAD=4.6 kcal/mol.[28] On the other hand, XYG3 was developed with 6-311 +G(3df,2p).[28] As the functional contains empirical parameters, the fitting procedure would have led to total errors that combine the intrinsic errors of the exchange-correlation functional and errors due to incompleteness of the basis set.

In the present work, we investigated the basis set dependence of XYG3, as well as those of B3LYP and MP2, in prediction of (1) HOFs, (2) bond dissociation enthalpies (BDEs), (3) reaction barrier heights (RBHs), and (4) nonbonded interactions (NBIs). We examined how well the XYG3 functional optimized at 6-311+G(3df,2p) could be transferred to smaller basis sets for practical use, and we developed the XYG3o functionals that were specifically optimized for smaller basis sets. We also performed Petersson's complete basis set (CBS) extrapolation[43] on the MP2-like correlation energy to see how XYG3/CBS would behave in completing the four tasks we considered here.

II. COMPUTATIONAL DETAILS

B3LYP is one of the first hybrid functionals that introduce a certain proportion of exact exchange E_x^{exact} based on the adiabatic connection formalism.[11,19,31]

$$E_{xc}^{\text{B3LYP}}[\rho] = E_{xc}^{\text{SVWN}} + c_1(E_x^{\text{exact}} - E_x^S) + c_2\Delta E_x^B + c_3\Delta E_c^{\text{LYP}}. \tag{1}$$

Here ΔE_x^B (Ref. 6) and ΔE_c^{LYP} (Ref. 7) are GGA correction terms to the LDA exchange-correlation SVWN.[3,4] The three mixing parameters $\{c_1,c_2,c_3\}=\{0.20,0.72,0.81\}$ were determined empirically against selected experimental thermochemical data of the G2/97 set.[41]

Recently, we developed the XYG3 functional[28] based on the adiabatic connection formalism[11,19,31] and the Görling–Levy coupling-constant perturbation expansion to the second order (PT2).[44] XYG3 is a doubly hybrid functional that mixes not only some exact exchange into E_x^{DFT} but also a certain portion of E_c^{PT2} into E_c^{DFT}.

$$E_{xc}^{\text{XYG3}}[\rho] = E_{xc}^{\text{SVWN}} + d_1(E_x^{\text{exact}} - E_x^S) + d_2\Delta E_x^B + d_3(E_c^{\text{PT2}} - E_c^{\text{LYP}}) + \Delta E_c^{\text{LYP}}. \tag{2}$$

XYG3 neglects the singles contribution as it did in the well known B2PLYP functional[24] and approximate E_c^{PT2} in a MP2-like fashion,

$$E_c^{\text{PT2}} \approx \frac{1}{4}\sum_{ij}\sum_{\alpha\beta}\frac{|\langle\varphi_i\varphi_j|\hat{v}_{ee}|\varphi_\alpha\varphi_\beta\rangle|^2}{\varepsilon_i+\varepsilon_j-\varepsilon_\alpha-\varepsilon_\beta}. \tag{3}$$

Here $\hat{v}_{ee}$ is the electron-electron repulsion operator, and the subscripts (i,j) and (α,β) denote the occupied and unoccupied KS orbitals, respectively. By fitting HOFs of the G3/99 set calculated with 6-311+G(3df,2p) to the experimental data,[41] the parameters in XYG3 are determined empirically as $\{d_1=0.8033,\ d_2=0.2107,\ d_3=0.3211\}$.[28]

We investigated here the basis set dependence for XYG3 prediction of HOFs, BDEs, RBHs, and NBIs. Each term of XYG3 in Eq. (2) at a given basis set was evaluated by using the converged B3LYP orbitals from the same basis set. The XYG3 results were compared with those of B3LYP and MP2. The basis sets examined in the present work included[42,45] B1: 6-311+G(d,p), B2: 6-311+G(2d,p), B3: 6-311+G(2d,2p), B4: 6-311+G(3d,2p), B5: 6-311 +G(2df,p), B6: 6-311+G(2df,2p), B7: 6-311+G(3df), B8: 6-311+G(3df,p), B9: 6-311+G(3df,2p), and B10: 6-311+ +G(3df,3pd). This choice echoes the common wisdom in MO theory that a TZ basis set is relatively complete for moderate accuracy and the major source of errors in calculating chemical reaction energies such as HOFs comes from the incompleteness of the polarization functions.[46]

DFT methods, by construction, consider all (core and valence) dynamic electron correlations. Hence, all calculations reported here include the core electron correlation in evaluation of the MP2-like term to match its DFT part. Previously,[47] we found that the frozen core approximation and Truhlar's scaling-all-correlation[23] method can be adopted to speed-up the calculations with similar accuracy after introducing an extra scaling parameter for the MP2-like term using Pople-type basis sets.

Martin and co-workers noticed that Petersson's CBS extrapolation scheme[43] can be used to mitigate the basis set incompleteness in the MP2-like correlation energy evaluation.[25] We carried out here CBS calculations with XYG3 using the B9 basis set 6-311+G(3df,2p). The pair extrapolation parameter $N_{\min}$ was set to 10.[25,43,45] We evaluated the performance of XYG3/CBS and compared the results with those of conventional MP2/CBS.

We used the well-known G3/99 set[41] as validating set of HOFs. B3LYP/6-311+G(d,p) was employed to optimize the equilibrium geometry of each species. Analytical harmonic frequency was calculated at the same level to ensure that each geometry corresponded to a true local minimum and to give zero-point energy (ZPE), thermocorrections. The standard heats of formation $\Delta_f H_{298}^{\theta}$ were calculated in the same

manner as Curtiss *et al.*[41] by first subtracting the calculated atomization enthalpies, using a scaled ZPE (0.9877) for the molecule, from the known experimental heats of formation of the isolated atoms at 0 K and then adding the calculated thermocorrections ($H_{0\to 298\ K}$) for the molecule, as well as $H_{0\to 298\ K}$ for elements in their standard states from experiments. In all our HOF calculations, spin-orbit (SO) corrections were also included.[41]

We calculated BDE according to the enthalpy change of the following reaction in the gas phase (298 K and 1 atm):

$$X - Y(\mathrm{g}) = X\cdot(\mathrm{g}) + Y\cdot(\mathrm{g}),$$

$$\mathrm{BDE} = \Delta_r H^o_{298} = \Delta_f H^o_{298}(X\cdot) + \Delta_f H^o_{298}(Y\cdot) - \Delta_f H^o_{298}(X-Y). \quad (4)$$

Here we supplied the experimental or calculated HOF with the given method for each species. When X or Y happened to be an atom, we used the experimental HOF.[41] As errors in HOFs might accumulate or cancel out in BDE calculations, such that BDEs carry additional information that is important for the judgment of functional performance for "real" chemistry. The so-called BDE92/07 set consisted of 92 bond dissociation reactions, set up by 27 radicals and 76 molecules contained in the G3/99 set.[33]

Zhao and Truhlar's NHTBH38/04 and HTBH38/04 sets were used to examine the XYG3 basis set dependence on calculations of barrier heights.[21–23] HTBH38/04 included forward and reverse barrier heights for 19 hydrogen transfer (HT) reactions, while NHTBH38/04 contained six heavy-atom transfer (HAT) reactions, eight nucleophilic substitution (NS) reactions, and five unimolecular and association (UM) reactions. Geometries and reference energies were taken from the Truhlar DB website.[21–23]

Zhao and Truhlar's NCIE31/05 set was used to investigate the XYG3 basis set dependence on calculations of noncovalent interaction.[21–23] NCIE31/05 was made of six hydrogen bond (HB) complexes, seven charge-transfer (CT) complexes, six dipole interaction (DI) complexes, seven weak interaction (WI) complexes, and five π-π stacking (PPS) complexes. Geometries and reference energies were taken from the Truhlar DB website without modifications.[21–23] All our calculations were performed by using the GAUSSIAN 03 suite of programs.[45]

III. RESULTS AND DISCUSSION

A. Heats of formation

The G3/99 set[41] is one of the commonly used set for developing and validating new DFT functionals in describing covalent bonding in the main group molecules.[11,15–18,20] The 223 molecules in the G3/99 set were historically divided into three subsets, G2-1, G2-2, and G3-3 with increasing average molecular size.[41] In a pioneer work of Pople and co-workers,[34] the G2-1 set was used to evaluate the BLYP[6,7] performance on HOFs with basis sets of 6-31G(d), 6-31+G(d), 6-311+G(2df,p), and 6-311+G(3df,2p). They used MP2/6-31G(d) geometries and vibrational frequencies (scaling factor 0.8929) at HF/6-31G(d), and did BLYP in a postmanner using HF density. They obtained MADs=7.14, 7.03, 4.32, and 3.94 kcal/mol, respectively, for these four basis sets. The role of the basis set in DFT calculations has been more systematically studied by Boese, Martin, and Handy using a modified G3/99 set.[38] Two of their main conclusions were the following. (1) The Pople basis sets,[42,45] developed for Hartree–Fock, also exhibited low errors for HOF calculations, while Dunning's correlation consistent basis sets,[48] which were optimized at the CISD level, did not seem to be an optimal choice for DFT calculations despite having a larger number of basis functions. (2) A TZ quality basis set was likely to be sufficient for use in GGA and hybrid calculations. Using still larger basis sets was not useful, as the basis set truncation error was probably about an order of magnitude lower than the functional error itself. We present here the XYG3 HOFs of the G3/99 set, along with the B3LYP and MP2 results, with basis sets ranging from B1 to B10. The detailed results can be found in the supplementary material (Tables S1–S3) (Ref. 49) and the statistic data are summarized in Table I, which are depicted in Fig. 1.

From Table I, it is seen that B3LYP shows a mild basis set dependence for the G2-1 set from B1 (MAD=4.26 kcal/mol) to B10 (MAD=2.19 kcal/mol). This basis set effect is enlarged significantly as the molecular size is increased. Thus for the G3-3 set, MAD associated with B3LYP/B1 is 22.20, and that from B3LYP/B10 is 7.57 kcal/mol. While the maximum positive error remains quite constant with increasing basis set, the maximum negative error is greatly reduced as the basis set is enlarged from B1 (max−=−88.30 kcal/mol) to B10 (max−=−19.22 kcal/mol). This might be expected as adding more polarization functions improves the description of the molecule, reducing the B3LYP tendency of underestimating the stability of the molecules.

Table I clearly shows that MP2 is more basis set dependent than B3LYP. Even for the G2-1 set, the MP2 MAD spans a range of 8.13 kcal/mol from B1 to B10, as opposed to the B3LYP range of 2.07 kcal/mol. Size dependence is also more severe in MP2. For the G3-3 set, the MP2 MAD range from B1 to B10 is increased to 37.24 kcal/mol, while that for B3LYP it is 14.63 kcal/mol. Similar to that in B3LYP, improving basis set reduces the MP2 max−. On the contrary, max+ increases significantly as the basis set is increased. Inspection on the data summarized in Table S2 (Ref. 49) shows that MP2 is most problematic for halogenated compounds, aromatic compounds, and hypervalent compounds, where their stabilities are underestimated with B1 by more than 20 kcal/mol, but overestimated with B10 by more than 20 kcal/mol.

In all cases, XYG3 is much less basis set dependent and size dependent than MP2 (see Table I and Fig. 1). When augmented with suitable number of polarization functions, XYG3 behaves significantly better than B3LYP. XYG3 has inherited both from B3LYP and MP2 some drawbacks, such that max+ and max− may occur at BeH and SF_6, respectively, as they do in B3LYP; while max+ increases as the basis set increases as it does in MP2, albeit at a much re-

TABLE I. Heats of formation at 298 K for the G3/99 Set (223). Errors [errors (Expt.-Theor.) are in kcal/mol; experimental data are from Ref. 41] for B3LYP, MP2, and XYG3 with different basis sets. [Codes for basis sets are B1: 6-311+G(d,p), B2: 6-311+G(2d,p), B3: 6-311+G(2d,2p), B4: 6-311+G(3d,2p), B5: 6-311+G(2df,p), B6: 6-311+G(2df,2p), B7: 6-311+G(3df), B8: 6-311+G(3df,p), B9: 6-311+G(3df,2p), and B10: 6-311++G(3df,3pd). CBS: the PT2 correlation energies are extrapolated from B9 to the complete basis set limit according to Petersson's CBS scheme using $N_{min}=10$ (Ref. 43).]

	MAD[a]					
Methods	G2-1	G2-2	G3-3	G3	max+[b]	max−[c]
B3LYP/B1	4.26	9.85	22.20	12.63	7.71 (BeH)	−88.30 (SF_6)
B3LYP/B2	2.75	6.59	15.77	8.73	7.87 (BeH)	−50.38 (SF_6)
B3LYP/B3	2.64	5.67	13.78	7.65	7.87 (BeH)	−50.38 (SF_6)
B3LYP/B4	2.29	4.67	12.22	6.62	7.88 (BeH)	−33.40 (SF_6)
B3LYP/B5	2.27	4.53	11.33	6.26	7.95 (BeH)	−30.37 (SF_6)
B3LYP/B6	2.20	3.92	9.56	5.39	7.98 (BeH)	−30.37 (SF_6)
B3LYP/B7	2.53	5.94	14.12	7.85	7.73 (BeH)	−30.80 (n-octane)
B3LYP/B8	2.20	3.93	9.76	5.46	7.97 (BeH)	−19.61 (n-octane)
B3LYP/B9	2.16	3.46	8.20	4.74	8.03 (BeH)	−19.22 (SF_6)
B3LYP/B10	2.19	3.32	7.57	4.47	8.08 (BeH)	−19.22 (SF_6)
UMP2/B1	13.96	26.25	50.13	31.25	3.01 (BCl_3)	−96.91 (n-octane)
UMP2/B2	11.25	19.77	36.56	23.32	8.55 (BF_3)	−83.67 (n-octane)
UMP2/B3	9.64	16.07	29.58	19.03	8.55 (BF_3)	−66.23 (n-octane)
UMP2/B4	8.91	14.40	25.35	16.73	18.62 (C_2F_6)	−61.30 (n-octane)
UMP2/B5	8.44	12.91	17.65	13.40	37.49 (C_2F_6)	−50.50 (n-octane)
UMP2/B6	7.13	10.43	12.94	10.46	37.49 (C_2F_6)	−31.55 ($Si(CH_3)_4$)
UMP2/B7	13.45	21.11	30.85	22.50	48.34 (C_2F_6)	−93.15 (n-octane)
UMP2/B8	8.32	12.90	16.33	12.92	48.34 (C_2F_6)	−41.86 (n-octane)
UMP2/B9	7.34	11.12	13.33	10.93	48.34 (C_2F_6)	−29.21 ($Si(CH_3)_4$)
UMP2/B10	5.83	10.28	12.89	10.06	48.34 (C_2F_6)	−16.07 (CN)
UMP2/CBS	7.69	17.00	26.64	17.95	77.36 (azulene)	−22.98 (Si_2H_6)
XYG3/B1	7.58	16.77	29.93	18.93	3.13 (BeH)	−106.05 (SF_6)
XYG3/B2	5.10	11.96	20.83	13.25	3.51 (BeH)	−60.59 (SF_6)
XYG3/B3	4.24	9.86	16.20	10.84	3.84 (BeH)	−60.59 (SF_6)
XYG3/B4	3.40	7.76	14.14	8.83	3.83 (BeH)	−40.23 (SF_6)
XYG3/B5	2.44	4.56	7.25	4.95	5.51 ($AlCl_3$)	−29.05 (SF_6)
XYG3/B6	1.76	2.59	3.71	2.76	6.26 (C_6H_5)	−29.14 (SF_6)
XYG3/B7	4.71	8.51	14.11	9.46	6.28 (BCl_3)	−34.45 (n-octane)
XYG3/B8	1.94	2.96	4.30	3.16	6.28 (BCl_3)	−16.62 (SF_6)
XYG3/B9	1.53	1.78	2.06	1.81	6.28 (BCl_3)	−16.67 (SF_6)
XYG3/B10	1.49	1.86	2.16	1.87	7.11 (naphthalene)	−16.57 (SF_6)
XYG3/CBS	2.21	4.34	7.18	4.77	17.40 (naphthalene)	−13.95 (SF_6)

[a]Mean absolute deviations.
[b]Maximum positive deviations.
[c]Maximum negative deviations.

duced rate. Overall, XYG3 presents a significant improvement over both B3LYP and MP2 in HOF predictions with B6, B9, and B10 basis sets.

Basis set extrapolation techniques have been applied to HOF predictions at various levels including HF,[50,51] MP2,[52,53] CCSD,[52,53] and CCSD(T),[54] as well as DFT.[35,36] They provided insights in tracing down the intrinsic errors of a specific method. These were generally done using either correlation consistent basis sets[48] or polarization consistent basis sets[36] with some dependence on the highest angular momentum included in the basis sets. These are generally expensive calculations and thus can only be applied to limited size of molecules for benchmarking. Petersson's CBS procedure is unique in its use of nonlinear pair natural orbital extrapolations to the CBS limit of the MP2 correlation energy.[43] The use of Petersson's CBS extrapolation suggests that smaller basis sets can be used, which can considerably speed the calculations and widen its range of applicability. We performed conventional UMP2/B9 calculations with all electrons being correlated and did CBS extrapolations. We did not see that such CBS calculations improved the UMP2 agreement with the experimental HOFs. This should be expected as MP2 is incomplete without taking into account of higher order correlation effects. On the other hand, we noticed that Ten-no and co-workers[53] gave MP2 correlation energy for benzene as −1.044 716 a.u. at aug-cc-pV6Z and −1.049 446 a.u. at aug-cc-pV7Z, with their best estimate as −1.0575(5) a.u., while our CBS calculation led to MP2 correlation energy of −1.200 356 a.u., overshooting Ten-no's best number by −89.64 kcal/mol. Indeed, our UMP2/CBS overestimated benzene's stability by 39.51 kcal/mol (see Table S4 for details).[49]

Previously, we found that B2PLYP degraded as the basis set size was reduced from CQZV3P (MAD=2.5 kcal/mol)

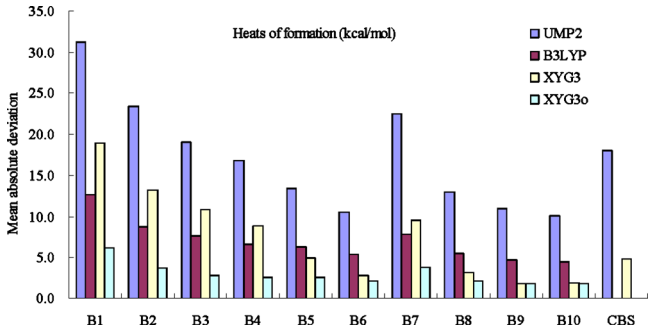

FIG. 1. Basis set dependence for the calculated heats of formation against the experimental data of the G3/99 set.

(Ref. 24) to Pople's 6-311+G(3df,2p) (MAD =4.6 kcal/mol).[28] Martin and co-workers found that the performance of B2PLYP deteriorated for basis sets larger than aug-pc2 and at Petersson's CBS limit, while their B2K-PLYP satisfyingly improved with improving the quality of basis set.[25] Table I and Fig. 1 show that the XYG3 performance generally improves with basis set from B1 to B10. However, XYG3/CBS does not improve the HOF results further. Besides the intrinsic errors in XYG3, we believe that the XYG3 compatibility with Petersson's CBS scheme may need more study (e.g., the proper choice of the pair extrapolation parameter N_{min} [25,43] for XYG3).

We performed some XYG3 calculations with correlation consistent basis sets.[48] We found that cc-pVDZ gave MAD of 28.69 kcal/mol, while cc-pVTZ reduced it to 3.63 kcal/mol for HOFs the G3/99 set. This is, however, inferior to the Pople basis sets of similar quality (e.g., B9 and B10), supporting Boese, Martin, and Handy's finding[38] that cc- basis sets did not seem to be an optimal choice for DFT calculations. It might be expected that even larger cc- basis set, in conjunction with basis set extrapolation, can lead to better results. In practice, we recommend to use XYG3 along with Pople's basis sets for effective, yet accurate, HOF predictions.[28,47]

B. Bond dissociation enthalpies

HOF based on atomization energy, where all chemical bonds are broken all at once, is a stringent test of the quality of a specific method. BDE, on the other hand, provides a more relevant test for ordinary chemistry where a specified A-B bond is broken, and it is generally expected that its statistic errors are reduced due to error cancellation in describing the parent molecules and the resultant radical species. Using the BDE92/07 set,[33] we have investigated the basis set dependence of XYG3 calculated BDEs, along with those from B3LYP and MP2, with basis sets ranging from B1 to B10. The detailed results are summarized in supplementary material (Table S5–S7).[49] The statistic data are presented in Table II and the trends are depicted in Fig. 2.

As it is clearly seen, basis set dependence is now not significant. For B3LYP, MAD from B1 is 5.86 kcal/mol and that from B10 is 5.18 kcal/mol. The most problematic case in the B3LYP BDE calculations is when the carbon atom is highly alkylated.[33,55] For example, B3LYP/B1 led to an error of 13.99 kcal/mol for $(CH_3)_3C-OCH_3$. Increasing basis set size does not help, such that B3LYP/B10 still led to an error of 13.24 kcal/mol for this C–O bond. There are only a few cases where improving basis set from B1 to B10 helps to an appreciable amount. Most of them happen for the carbon-halogen bond. For example, B3LYP/B1 gave an error of 8.23 kcal/mol for $CH_3C(O)-Cl$, whereas B3LYP/B10 reduced it to 2.48 kcal/mol (see Table S5 for more examples).[49]

Basis set dependence of the UMP2 BDEs is also significantly attenuated as compared to that of the UMP2 HOFs, while it is still sizable as compared to that of the B3LYP BDEs. MAD from UMP2/B1 is 6.18 kcal/mol, which increases, rather than decreases as would have been expected, to 8.41 for UMP2/B10 and further increases to 9.97 if CBS is applied (see Table II). Increasing basis set size did help in many cases as shown by data in Table S6.[49] For example, UMP2/B1 led to an error of 6.20 kcal/mol for H–NH, which was reduced to 0.52 kcal/mol with UMP2/B10. However, there are many other cases that larger basis sets actually worsen the UMP2 agreement with the experiments. If the specified bond is near a double bond or a triple bond, the UMP2 results are poor, which are made even worse when the basis set size is increased. For example, an error of −45.75 kcal/mol was resulted from the UMP2/B1 calculation for NC–CN, which increased to −51.73 (or −55.60) kcal/mol if UMP2/B10 (or UMP2/CBS) was used (see Table S6 for more examples).[49]

Table II and Table S7 (Ref. 49) show that XYG3 is particularly satisfying for BDE predictions. Even with XYG3/B1, MAD for BDE92/07 is 3.29 kcal/mol, smaller than the best values of UMP2/B1 (6.18) and B3LYP/B10 (5.17). Improving basis set improves steadily the XYG3 performance, such that XYG3/B10 and XYG3/CBS give MADs=1.46 and 1.32 kcal/mol, respectively. Previously we showed that G2 and G3 led to MADs=1.8 and 1.1 kcal/mol, respectively, for this same BDE set.[33] The severe underestimation of B3LYP for alkylated C-X bond is largely corrected in XYG3, giving an error of 2.62 kcal/mol for $(CH_3)_3C-OCH_3$ with B10. The dramatic overestimation of UMP2 for NC–CN is removed in XYG3, which shows an error of only 0.43 kcal/mol with B10 (see Table S7 for more examples).[49]

C. Reaction barrier heights

DFT methods usually underestimate RBHs.[19,21–23,28,47] Such errors are usually attributed to the self-interaction errors (SIE) that make local DFT functionals problematic for the stretched partially broken bonds, characteristic of the transition states for chemical reactions.[19] Table III and Fig. 3 display the statistic behaviors for B3LYP, UMP2, and XYG3, respectively, for RBH estimations with various basis sets. The corresponding details are summarized in Tables S9–S14 for B3LYP, Tables S15–S20 for UMP2, and Tables S21–S26 for XYG3.[49] The CBS results with UMP2 and XYG3 are listed in Tables S27–S32.[49]

Again the basis set dependence is quite mild for B3LYP. This is also true for XYG3, showing that TZ basis set plus a minimum set of polarization functions is generally good for barrier height predictions. An MAD of 4.6 kcal/mol is encountered for B3LYP with various basis sets for a total of 76

TABLE II. Bond dissociation enthalpies at 298 K (92). Errors (MADs are in kcal/mol; experimental data are from Refs. 41 and 33.) for B3LYP, MP2, and XYG3 with different basis sets [codes for basis sets are B1: 6-311+G(d,p), B2: 6-311+G(2d,p), B3: 6-311+G(2d,2p), B4: 6-311+G(3d,2p), B5: 6-311+G(2df,p), B6: 6-311+G(2df,2p), B7: 6-311+G(3df), B8: 6-311+G(3df,p), B9: 6-311+G(3df,2p), and B10: 6-311++G(3df,3pd); CBS: the PT2 correlation energies are extrapolated from B9 to the complete basis set limit according to Petersson's CBS scheme using $N_{min}=10$ (Ref. 43)] for various bond types.

Method	C–H (15)	X-H (10)[a]	C–C (22)	C–O (16)	C–N (6)	C–F (4)	C–Cl (7)	C–S (5)	X-Y (7)[b]	Total MAD
B3LYP/B1	2.53	3.09	6.55	8.00	7.34	6.53	6.31	7.85	6.30	5.85
B3LYP/B2	2.50	3.11	6.47	7.64	7.71	5.15	5.26	7.28	5.30	5.55
B3LYP/B3	2.43	2.77	6.50	7.65	7.83	5.11	5.20	7.27	5.36	5.51
B3LYP/B4	2.45	2.71	6.41	7.35	7.84	4.34	4.46	6.67	5.18	5.30
B3LYP/B5	2.63	3.28	6.51	7.19	7.34	4.00	4.88	6.75	4.76	5.35
B3LYP/B6	2.46	2.94	6.48	7.11	7.23	3.58	4.63	6.76	4.84	5.22
B3LYP/B7	3.04	4.23	6.17	6.80	7.37	3.16	4.23	6.25	4.63	5.25
B3LYP/B8	2.68	3.13	6.52	7.08	7.60	3.19	4.35	6.48	4.75	5.25
B3LYP/B9	2.61	2.91	6.58	7.08	7.56	3.17	4.33	6.51	4.75	5.22
B3LYP/B10	2.58	2.76	6.54	7.07	7.53	3.12	4.20	6.44	4.75	5.17
UMP2/B1	6.14	4.72	9.70	4.63	5.74	7.52	5.66	2.51	3.59	6.18
UMP2/B2	5.99	4.44	10.09	5.77	5.80	9.20	6.51	1.91	2.66	6.45
UMP2/B3	5.72	4.12	10.05	5.89	5.81	9.28	6.76	1.84	2.66	6.41
UMP2/B4	5.76	3.95	10.36	6.40	6.35	10.54	8.01	1.84	2.77	6.75
UMP2/B5	6.06	3.95	12.29	8.06	8.57	13.38	10.20	2.62	3.84	8.11
UMP2/B6	5.70	3.63	12.36	8.20	8.57	13.45	10.41	2.82	3.86	8.09
UMP2/B7	7.55	6.39	13.18	9.17	9.73	14.90	11.24	3.29	4.21	9.31
UMP2/B8	6.01	3.63	12.58	8.66	8.93	14.77	11.02	3.07	3.96	8.42
UMP2/B9	5.70	3.44	12.42	8.71	9.07	14.80	11.14	3.16	4.14	8.36
UMP2/B10	5.31	3.19	12.72	8.82	9.08	14.87	11.54	3.68	4.01	8.41
UMP2/CBS	5.46	4.01	14.65	10.66	11.41	17.60	12.72	5.69	6.62	9.97
XYG3/B1	1.36	2.49	2.64	4.07	3.30	7.43	3.53	3.61	5.93	3.29
XYG3/B2	1.25	2.47	2.52	3.42	3.46	5.16	2.50	3.45	4.34	2.83
XYG3/B3	0.91	1.63	2.54	3.40	3.57	5.10	2.43	3.35	4.34	2.68
XYG3/B4	0.96	1.54	2.29	2.84	3.40	4.15	1.69	2.67	3.95	2.34
XYG3/B5	1.25	2.38	1.79	1.79	2.18	2.55	1.02	1.70	2.53	1.82
XYG3/B6	1.22	1.63	1.97	2.01	2.97	3.48	1.57	1.63	2.54	1.94
XYG3/B7	2.52	4.45	1.26	0.95	1.72	1.49	0.96	1.32	2.17	1.85
XYG3/B8	1.24	2.02	1.70	1.40	2.25	1.54	0.85	1.56	2.41	1.62
XYG3/B9	1.02	1.62	1.80	1.39	2.20	1.50	0.86	1.55	2.39	1.56
XYG3/B10	0.94	1.27	1.68	1.33	2.17	1.44	1.03	1.43	2.36	1.46
XYG3/CBS	0.85	1.41	1.60	1.02	1.76	1.10	0.94	1.00	2.41	1.32

[a]X-H: N–H, O–H, Si–H, P–H, and S–H.
[b]X-Y: O–O, N–N, N–O, C–Si, Si–Si, Cl–O, and Cl–N.

barrier heights (see Table III). A uniformed tendency for B3LYP to underestimate RBHs is evident for the forward reaction in Truhlar's NHTBH38/04 set (see Table S10).[49] The worst happens at the subset of HAT reactions, whose MAD is as high as 8.5 kcal/mol. XYG3 is obviously superior to B3LYP, leading to MAD around 1.0 kcal/mol (see Table III). This is probably due to its large portion (~80%) of the HF-like exchange that reduces the SIE. Increasing the basis set size has the most profound effect on reducing the XYG3 errors for the HAT12 subset. MAD (3.33 kcal/mol) associated with XYG3/B1 is reduced to 1.42 for XYG3/B10 and further to 1.15 for XYG3/CBS.

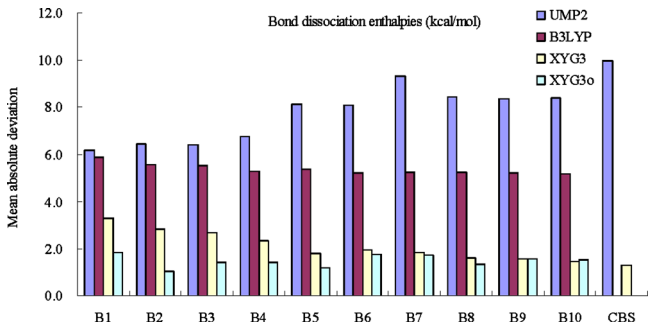

FIG. 2. Basis set dependence for the calculated bond dissociation enthalpies against the experimental data of the BDE92/07 set.

As it is clearly seen from Table III and Fig. 3, basis set dependence for UMP2 is now not that dramatic. MAD from B1 is 5.82 kcal/mol and that from B10 is 4.44 kcal/mol. UMP2/CBS further reduced MAD to 4.04 kcal/mol. Hence, improving basis set quality indeed improves the agreement with the reference data. The most problematic cases in the UMP2 barrier calculations are HAT reactions, whose MAD is generally more than 10 kcal/mol. UMP2 is however quite satisfactory for the set of NS reactions.

Previously, we have tested some DFT methods, along with HF and MP2, for describing the whole $H+CH_4 \rightarrow H_2$

TABLE III. Reaction barrier heights for Truhlar's NHTBH38/04 set (UM10: ten data points for association and unimolecular reactions, NS16: 16 data points for nucleophilic substitution reactions, HAT12: 12 data points for heavy-atom transfer reactions) and HTBH38/04 set (HT38: 38 data points for hydrogen transfer reactions). Errors (MADs are in kcal/mol; W1 reference data are from Refs. 21–23) for B3LYP, MP2, and XYG3 with different basis sets. [Basis set codes are B1: 6-311+G(d,p), B2: 6-311+G(2d,p), B3: 6-311+G(2d,2p), B4: 6-311+G(3d,2p), B5: 6-311+G(2df,p), B6: 6-311+G(2df,2p), B7: 6-311+G(3df), B8: 6-311+G(3df,p), B9: 6-311+G(3df,2p), and B10: 6-311++G(3df,3pd). CBS: the PT2 correlation energies are extrapolated from B9 to the complete basis set limit according to Petersson's CBS scheme using $N_{min}=10$ (Ref. 43).]

Method	UM10	NS16	HAT12	HT38	Total (76)
B3LYP/B1	2.00	3.89	8.91	4.46	4.72
B3LYP/B2	1.99	4.32	8.71	4.45	4.77
B3LYP/B3	2.03	4.34	8.48	4.42	4.73
B3LYP/B4	2.09	4.13	8.47	4.46	4.71
B3LYP/B5	1.97	3.49	8.72	4.45	4.60
B3LYP/B6	2.02	3.48	8.55	4.42	4.56
B3LYP/B7	1.88	3.56	8.43	3.97	4.32
B3LYP/B8	1.97	3.35	8.51	4.47	4.55
B3LYP/B9	2.02	3.38	8.51	4.43	4.54
B3LYP/B10	2.04	3.27	8.82	4.56	4.63
UMP2/B1	6.00	3.52	12.83	4.53	5.82
UMP2/B2	5.69	0.68	11.77	4.19	4.85
UMP2/B3	5.60	0.66	12.07	4.08	4.82
UMP2/B4	5.41	0.85	11.73	3.91	4.70
UMP2/B5	5.66	1.21	11.39	3.94	4.77
UMP2/B6	5.51	0.98	11.60	3.83	4.68
UMP2/B7	6.18	1.41	12.24	5.58	5.83
UMP2/B8	5.54	2.09	11.34	3.86	4.89
UMP2/B9	5.43	1.69	11.42	3.82	4.78
UMP2/B10	5.06	1.77	10.83	3.38	4.44
UMP2/CBS	5.07	1.60	9.47	3.09	4.04
XYG3/B1	1.23	0.85	3.33	1.14	1.35
XYG3/B2	1.00	2.66	2.12	1.05	1.50
XYG3/B3	1.02	2.81	2.42	0.96	1.53
XYG3/B4	1.03	2.50	2.16	0.85	1.37
XYG3/B5	0.99	1.66	1.77	0.91	1.17
XYG3/B6	1.06	1.80	1.67	0.84	1.16
XYG3/B7	1.15	1.59	2.98	1.49	1.63
XYG3/B8	0.91	1.21	1.80	0.70	0.96
XYG3/B9	0.98	1.42	1.66	0.75	1.02
XYG3/B10	0.83	1.36	1.42	0.68	0.92
XYG3/CBS	0.82	1.39	1.15	0.84	1.00

$+CH_3$ reaction path using CCSD(T)/6-311++G(3df,2pd) data as the reference.[28,47] We showed that the XYG3 results are within 0.44 kcal/mol of the CCSD(T) results for the entire reaction path. On the other hand, HF and MP2 are both inadequate for potential energy surface calculations. HF overestimates the barrier height by 8.35 kcal/mol and MP2 overestimates the barrier height and endothermicity of the reaction by 5.81 and 4.70 kcal/mol, respectively. We may view HF and MP2 as a kind of functionals with 100% HF exchange plus null and 100% MP2 correlation, respectively. In contrast with the B3LYP results, a uniformed tendency for UMP2 to overestimate RBHs is clearly evident for the forward reaction in Truhlar's NHTBH38/04 set (see Table S16).[49] This infers that besides the exchange functional, the correlation functional is also critical for an adequate description of RBHs.

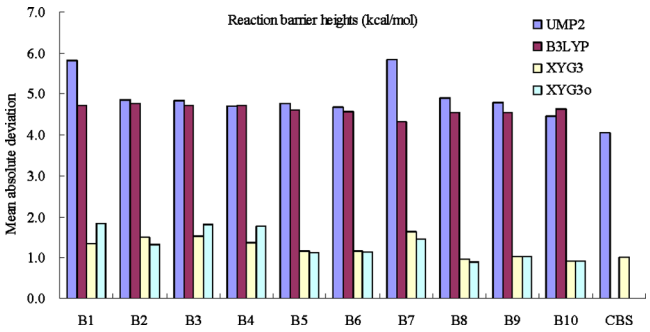

FIG. 3. Basis set dependence for the calculated reaction barrier heights against the benchmark data of the NHTBH38/04 and HTBH38/04 sets.

D. Noncovalent interactions

Despite that DFT is now the leading first principles method for applications in physics, chemistry, biology, and materials science, a known serious deficiency for traditional DFT methods lies in its poor behavior in describing NBIs, which are so important to the packing of molecules into sol-

TABLE IV. Noncovalent interaction energies for Truhlar's NCIE31/05 (NCIE31/05 set consists of six HB complexes, seven CT complexes, six DI complexes, seven WI complexes, and five PPS complexes) set. Errors (MADs are in kcal/mol; W1 reference data are from Refs. 21–23) for B3LYP, MP2, and XYG3 with different basis sets. [Basis set used for single point calculations. B1: 6-311+G(d,p), B2: 6-311+G(2d,p), B3: 6-311+G(2d,2p), B4: 6-311+G(3d,2p), B5: 6-311+G(2df,p), B6: 6-311+G(2df,2p), B7: 6-311+G(3df), B8: 6-311+G(3df,p), B9: 6-311+G(3df,2p), and B10: 6-311++G(3df,3pd). CBS: the PT2 correlation energies are extrapolated from B9 to the complete basis set limit according to Petersson's CBS scheme using $N_{min}=10$ (Ref. 43).]

Method	HB6	CT7	DI6	WI7	PPS5	Total (31)
B3LYP/B1	0.88	1.49	0.55	0.30	2.97	1.16
B3LYP/B2	0.73	1.28	0.53	0.30	2.98	1.08
B3LYP/B3	0.59	1.14	0.56	0.30	3.05	1.04
B3LYP/B4	0.61	1.08	0.53	0.28	2.97	1.01
B3LYP/B5	0.70	0.93	0.61	0.29	2.92	1.00
B3LYP/B6	0.60	0.77	0.61	0.29	2.98	0.96
B3LYP/B7	0.67	0.82	0.53	0.27	2.89	0.94
B3LYP/B8	0.61	0.78	0.58	0.28	2.92	0.94
B3LYP/B9	0.63	0.76	0.61	0.27	2.93	0.95
B3LYP/B10	0.60	0.79	0.66	0.28	2.88	0.95
UMP2/B1	0.97	0.53	0.24	0.08	1.85	0.67
UMP2/B2	0.58	1.10	0.26	0.08	1.50	0.67
UMP2/B3	0.28	0.96	0.49	0.07	1.48	0.62
UMP2/B4	0.35	1.36	0.63	0.13	1.83	0.82
UMP2/B5	0.51	0.83	0.29	0.09	1.41	0.59
UMP2/B6	0.36	0.65	0.60	0.08	1.38	0.57
UMP2/B7	0.34	1.28	0.64	0.15	1.79	0.80
UMP2/B8	0.25	1.13	0.55	0.13	1.72	0.72
UMP2/B9	0.33	1.15	0.68	0.14	1.66	0.75
UMP2/B10	0.53	1.16	0.94	0.18	1.93	0.90
UMP2/CBS	0.35	1.70	1.75	1.14	···	1.18
XYG3/B1	0.85	0.88	0.24	0.08	0.23	0.47
XYG3/B2	0.59	0.88	0.20	0.08	0.32	0.42
XYG3/B3	0.39	0.82	0.20	0.08	0.39	0.38
XYG3/B4	0.39	0.87	0.22	0.12	0.25	0.38
XYG3/B5	0.62	0.66	0.19	0.09	0.30	0.37
XYG3/B6	0.42	0.58	0.19	0.09	0.37	0.33
XYG3/B7	0.41	0.70	0.23	0.10	0.18	0.33
XYG3/B8	0.36	0.64	0.17	0.11	0.22	0.31
XYG3/B9	0.38	0.64	0.19	0.12	0.26	0.32
XYG3/B10	0.46	0.62	0.27	0.13	0.20	0.34
XYG3/CBS	0.35	0.58	0.59	0.41	···	0.46

ids, the binding of drug molecules to proteins. Encouragingly, important achievements along this line have been made recently.[21–28,56] Here we test the basis set dependence of XYG3, as well as B3LYP and UMP2, using Truhlar's NCIE31/05 set.[21–23] The statistic errors are summarized in Table IV with details given in Table S33–S36.[49]

B3LYP/B10 led to MAD of 0.95 kcal/mol for NCIE31/05, which is an improvement over B3LYP/B1 (MAD =1.16 kcal/mol). With B1, HB strength may be overestimated due to basis set superposition error (BSSE),[57] while with B10, it is clear that B3LYP has a tendency to underestimate HB strength [e.g., $(NH_3)_2$, $(H_2O)_2$, and $(NH_3)(H_2O)$ in Table S33].[49] Basis set effects are quite significant in the CT complexes (see Table IV). B3LYP/B1 resulted in a MAD of 1.49 kcal/mol, which was reduced to 0.79 with B10. Generally, B3LYP overestimates the interaction strength in the CT complexes. The $(NH_3)(F_2)$ complex provides an extreme example. While B3LYP/B1 is overbinding by more than 100%, it is still so by 59% with the B10 basis set. Basis set dependence of B3LYP is, however, mild for the other three sets (DI, WI, and PPS), showing a general trend for underbinding. Obviously, B3LYP is particularly poor for the PPS subset for which London dispersion dominates. Errors can be more than 200%, erroneously showing, for example, that benzene dimer is unbound (see Table S33 for more details).[49]

UMP2 is physically sound in describing dispersion interaction by design.[58] In practice, however, the overbinding tendency is still dramatic, even when basis set is as large as B10, if BSSE is not properly corrected (see Table S34).[49] Table IV shows that UMP2/B1 led to MAD=0.67, which actually increased to 0.90 kcal/mol with B10 (see Table IV).

There are many detailed basis set extrapolation studies using correlation consistent basis sets published in the literature,[58–60] which may be compared with the results obtained here with Petersson's CBS procedure (Table S36).[49] For example, Boese and co-workers reported a BSSE corrected MP2-R12/aug-cc-pV5Z value for $(HCl)_2$ HB strength

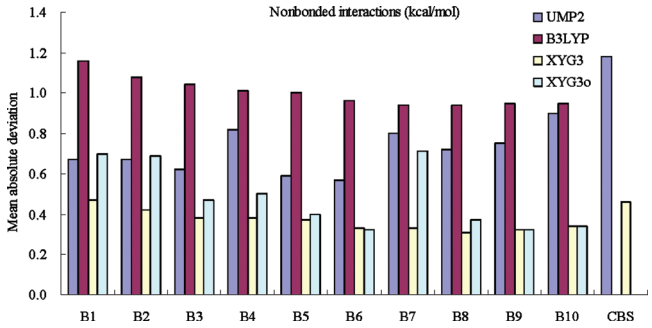

FIG. 4. Basis set dependence for the calculated nonbonded interactions against the benchmark data of the NCIE31/05 set.

of 2.24 kcal/mol, while a T-Q extrapolation led to 2.21 kcal/mol.[60] In comparison, our MP2/B9 gave a value of 2.46 kcal/mol, and Petersson's CBS led to 4.09 kcal/mol, being significantly too high. Table S36 displays several other cases of such failure [e.g., $(H_2O)(ClF)$, $(H_2S)(HCl)$, $(He)_2$, AND (He)(Ar)].[49] Petersson's CBS scheme relies on using proper localization procedure. Indeed, we encountered localization failure for benzene dimers, which limits its applicability in calculating NBIs.

Table IV demonstrates the XYG3 performance with various basis sets for the NCIE31/05 set. The basis set dependence is mild as shown in Fig. 4. While XYG3/B10 led to MAD=0.34 kcal/mol, MAD associated with XYG3/B1 is only 0.47 kcal/mol, being quite satisfactory. Improving the basis set quality improves the XYG3 performance for HB and CT complexes. Note that we did not include BSSE corrections in all our calculations. Previously, Sherrill and coworkers calculated the XYG3 potential energy curve for the $CH_4-C_6H_6$ complex.[61] They found that the XYG3 curve is essentially coincident with the CCSD(T) curve to the left of equilibrium and coincident with the B2PLYP-D curve to the right of equilibrium, whereas the counterpoise-corrected XYG3 curve is underbonded, being less accurate than the uncorrected XYG3.[61] It is suggested that BSSE corrections may not be applied to the XYG3 method[47,61] and it might be anticipated that adding a suitable dispersion term may further increase the XYG3 accuracy.

E. Basis set optimized functionals XYG3o

Scheiner and co-workers explored the basis set dependence of KS DFT using SVWN,[3,4] BLYP,[6,7] BPW91,[5,6] and B3PW91[5,6,11] with six basis sets [DZP,[62] 6-31G**,[63] DZVP,[64] TZVP,[64] TZ2P,[62] and uncontracted aug-cc-pVTZ (Ref. 48) (UCC)].[37] They observed that there was a significant improvement on reaction energetics for B3PW91 in increasing the basis set size, and suggested that there might be value in deriving functionals specifically tailored for each basis set. Boese and co-workers systematically explored the role of basis set in KS DFT.[38] They have experimented with basis set optimized GGA and hybrid GGA functionals and found that the functionals obtained by fitting to a TZ quality basis set are transferable to other basis sets. They therefore proposed that it might not be important to reach the basis set limit when developing new density functionals, since the overall DFT errors were considerably larger than basis set truncation errors.[38] Gill and Pople[65] would be the first to develop a functional specifically for small basis set for practical use. In their development of a GGA type of functional EDF1, a 6-31+G(d) basis set was employed. We also note that Truhlar's popular MPW1K (a hybrid GGA) was developed using 6-31+G(d,p).[66] As a doubly hybrid functional contains MP2-like correlation, it is particularly appealing to use smaller basis set for speedup, while it is not clear how well the functional parametrization at smaller basis set can absorb some deficiencies of the basis set itself. We have optimized XYG3 with B1–B10 basis sets using only HOFs from the G3/99 set as we did before in developing the original XYG3 (Ref. 28) and the recent XYG3s (Ref. 47) functionals. The resultant functionals are called XYG3o.

Table V summarizes the optimized parameters for each basis set optimized functional. The B7 derived functional would be of least practical value, as this is an unbalanced

TABLE V. Basis set optimized functionals XYG3o: functional parameters and functional performances.

	Parameters[b]					MAD[c]			
Method[a]	E_x^{exact}	E_x^S	ΔE_x^B	E_c^{LYP}	E_x^{PT2}	HOF	BDE	RBH	NBI
XYG3o/B1	0.9035	0.0965	0.1031	0.5179	0.4821	6.11	1.85	1.84	0.70
XYG3o/B2	0.8336	0.1664	0.1002	0.6701	0.3299	3.70	1.05	1.32	0.69
XYG3o/B3	0.8153	0.1847	0.1873	0.6115	0.3885	2.85	1.41	1.82	0.47
XYG3o/B4	0.8066	0.1934	0.2041	0.6178	0.3822	2.57	1.41	1.77	0.50
XYG3o/B5	0.8043	0.1957	0.1852	0.6699	0.3301	2.56	1.17	1.11	0.40
XYG3o/B6	0.8039	0.1961	0.2123	0.6649	0.3351	2.12	1.75	1.14	0.32
XYG3o/B7	0.7890	0.2110	0.1060	0.7440	0.2560	3.84	1.74	1.45	0.71
XYG3o/B8	0.7780	0.2220	0.2061	0.7004	0.2996	2.13	1.36	0.90	0.37
XYG3o/B9	0.8033	0.1967	0.2107	0.6789	0.3211	1.81	1.56	1.02	0.32
XYG3o/B10	0.8033	0.1967	0.2107	0.6825	0.3175	1.80	1.52	0.91	0.34

[a]Basis sets used for functional optimizations. B1: 6-311+G(d,p), B2: 6-311+G(2d,p), B3: 6-311+G(2d,2p), B4: 6-311+G(3d,2p), B5: 6-311+G(2df,p), B6: 6-311+G(2df,2p), B7: 6-311+G(3df), B8: 6-311+G(3df,p), B9: 6-311+G(3df,2p), and B10: 6-311++G(3df,3pd).
[b]The coefficients of the exact exchange and the Slater exchange are normalized to 1.0 and the coefficients of the LYP correlation and PT2 correlation are normalized to 1.0.
[c]MADs are in kcal/mol. HOF: heats of formation for the G3/99 set (Ref. 41). BDE: bond dissociation enthalpies for the BDE92/07 set (Refs. 41 and 33). RBH: reaction barrier heights for the NHTBH38/04 and HTBH38/04 sets (Refs. 21–23). NBI: nonbonded interactions for the NCIE31/05 set (Refs. 21–23).

basis set with no polarization function on hydrogen atom. Table V displays a trend that smaller basis sets favors higher ratios of exact exchange and PT2 correlation. This is most significant for B1. For all other XYG3o, the portion for exact exchange is around 80% and that for PT2 correlation is around 35%, being close to the original XYG3 functional.

Table V also presents the XYG3o performance, which is depicted in Figs. 1–4 to compare with that of XYG3 before optimization. The detailed results are tabulated in Tables S37–S45.[49] Optimization indeed absorbs the deficiencies of the basis set itself. This is particular so for HOF predictions with basis sets without including f function (B1–B4). MADs are reduced by 12.82, 9.55, 7.99, and 6.26 kcal/mol for B1, B2, B3, and B4, respectively. Note that XYG3o/B1 still gave MAD=6.11 kcal/mol for HOFs, reflecting the inherent poor quality of this basis set. For application purposes, BDE is of higher significance. Even though BDE is not included in the training set, we see that the BDE prediction is consistently improved with XYG3o. MADs for BDE predictions with B1–B10 are all below 2 kcal/mol, better than the situation for HOF predictions. By contrast, B3LYP with B9 and B10 gave MADs=4.74 and 4.47 kcal/mol, respectively, for HOFs, which worsen to 5.22 and 5.17, respectively, for BDEs. XYG3o also works fine for barrier height predictions, whose performance is similar to that of XYG3. XYG3o works less satisfactorily for NBIs, where functional optimization based only on HOFs may be biased against the nonbounded interaction complexes. Generally, we may recommend XYG3o/B5 or B6 as a cost-effective replacement of XYG3/B9 or B10 for larger systems.

IV. CONCLUSION

In this study, we addressed the basis set dependence issue associated with the newly developed doubly hybrid functional XYG3. The results were calibrated with either experimental data for HOFs and BDEs, or benchmark calculations for RBHs and NBIs. In addition to the XYG3 calculations, we also carried out the corresponding B3LYP and MP2 calculations. These two traditional DFT and *ab initio* methods might be considered as the parent methods upon which XYG3 was based. The comparison among XYG3, B3LYP, and MP2 will help to understand the origin of errors associated with XYG3. We performed Petersson's CBS calculations, trying to identify the intrinsic errors associated with the XYG3 functional. The main conclusions from this study are summarized as follows:

(1) In general, the basis set convergence behaviors for B3LYP and XYG3 methods are similar, since the major components of the energy are similar in the two cases. UMP2 results are more prone to basis set deficiencies, while XYG3 only contains 32% MP2-like correlation, which lessens its infection.
(2) HOFs are most sensitive to basis set quality. There is continued improvement with the larger basis sets for all three methods. The smallest MADs achieved by each method for the G3/99 set are 10.06 (UMP2), 4.47 (B3LYP), and 1.81 kcal/mol (XYG3).
(3) There is little basis set dependence for the B3LYP BDEs, whose MADs fall in the range of 5.85–5.17 kcal/mol. For XYG3, with the B1 basis set, MAD is 3.29 and improves continuously to 1.46 kcal/mol with the B10 basis set. For BDE predictions, none of the UMP2 results are satisfactory. MAD with B1 is 6.18 and degrades to 8.41 kcal/mol with B10, showing the inherent limitation of this method.
(4) Basis dependences for RBHs are all similar for UMP2, B3LYP, and XYG3, which improve steadily, but slightly, with larger basis sets. Within B1 and B10, MADs remain between 5.82 and 4.44 kcal/mol for UMP2, 4.72 and 4.63 for B3LYP, and 1.35 and 0.92 for XYG3.
(5) B3LYP is unable to bind the PPS complexes, while MP2 overbinds, and enlarging the basis set does not change this situation. XYG3 is the best performer for practical use with moderate size basis set for NBIs, partly due to BSSE.
(6) Petterson's CBS generally did not improve the XYG3 agreement with reference data. This, on one hand, suggests that basis set originally used in XYG3 parametrization becomes an integral part of the model, it, on the other hand, questions the compatibility of Petterson's CBS scheme with the XYG3 model, especially for NBIs.
(7) Fitting will to some extent compensate for the inadequacies in the basis set, leading to the XYG3o models specifically optimized for each basis set. The XYG3o results generally improve over XYG3 for a given basis set. The limit that a XYG3o functional can push reflects the inherent quality and compatibility of the basis set used in functional optimization. Generally, we recommend XYG3o/B5 or B6 as a cost-effective replacement of XYG3/B9 or B10 for larger systems.

ACKNOWLEDGMENTS

This work was supported by NSFC (Grant No. 10774126), the Ministry of Science and Technology (Grant No. 2007CB815206, 2011CB808504) and the Swedish National Infrastructure for Computing (SNIC).

[1] R. G. Parr and W. Yang, *Density Functional Theory of Atoms and Molecules* (Oxford University Press, New York, 1989); W. Koch and M. C. Holthausen, *A Chemist's Guide to Density Functional Theory* (Wiley-VCH, Berlin, 2001).
[2] P. Hohenberg and W. Kohn, Phys. Rev. B **136**, B864 (1964); W. Kohn and L. J. Sham, Phys. Rev. A **140**, A1133 (1965).
[3] P. A. M. Dirac, Proc. Cambridge Philos. Soc. **26**, 376 (1930); J. C. Slater, *Quantum Theory of Molecules and Solids* (McGraw-Hill, New York, 1974), Vol. 4.
[4] S. H. Vosko, L. Wilk, and M. Nusair, Can. J. Phys. **58**, 1200 (1980).
[5] J. P. Perdew and Y. Wang, Phys. Rev. B **45**, 13244 (1992).
[6] A. D. Becke, Phys. Rev. A **38**, 3098 (1988).
[7] C. Lee, W. Yang, and R. G. Parr, Phys. Rev. B **37**, 785 (1988).
[8] J. P. Perdew, K. Burke, and M. Ernzerhof, Phys. Rev. Lett. **77**, 3865 (1996); **78**, 1396(E) (1997).
[9] T. Van Voorhis and G. E. Scuseria, J. Chem. Phys. **109**, 400 (1998).
[10] J. M. Tao, J. P. Perdew, V. N. Staroverov, and G. E. Scuseria, Phys. Rev. Lett. **91**, 146401 (2003).
[11] A. D. Becke, J. Chem. Phys. **98**, 5648 (1993); **98**, 1372 (1993).
[12] A. D. Rabuck and G. E. Scuseria, Theor. Chem. Acc. **104**, 439 (2000);

Chem. Phys. Lett. **309**, 450 (1999); S. N. Maximoff and G. E. Scuseria, J. Chem. Phys. **114**, 10591 (2001).
[13] F. A. Hamprecht, A. J. Cohen, D. J. Tozer, and N. C. Handy, J. Chem. Phys. **109**, 6264 (1998); A. J. Cohen and N. C. Handy, Chem. Phys. Lett. **316**, 160 (2000); A. D. Boese, N. L. Doltsinis, N. C. Handy, and M. Sprik, J. Chem. Phys. **112**, 1670 (2000); A. D. Boese and N. C. Handy, *ibid.* **114**, 5497 (2001); W.-M. Hoe, A. J. Cohen, and N. C. Handy, Chem. Phys. Lett. **341**, 319 (2001).
[14] G. Menconi, P. J. Wilson, and D. J. Tozer, J. Chem. Phys. **114**, 3958 (2001); P. J. Wilson, T. J. Bradley, and D. J. Tozer, *ibid.* **115**, 9233 (2001).
[15] A. D. Becke, J. Chem. Phys. **107**, 8554 (1997); H. L. Schmider and A. D. Becke, *ibid.* **109**, 8188 (1998); A. D. Becke, J. Comput. Chem. **20**, 63 (1999).
[16] M. Filatov and W. Thiel, Phys. Rev. A **57**, 189 (1998); Mol. Phys. **91**, 847 (1997).
[17] C. Adamo and V. Barone, J. Chem. Phys. **108**, 664 (1998); **110**, 6158 (1999).
[18] A. D. Becke, J. Chem. Phys. **122**, 064101 (2005).
[19] P. Mori-Sánchez, A. J. Cohen, and W. Yang, J. Chem. Phys. **124**, 091102 (2006); A. J. Cohen, P. Mori-Sánchez, and W. Yang, *ibid.* **126**, 191109 (2007).
[20] X. Xu and W. A. Goddard III, Proc. Natl. Acad. Sci. U.S.A. **101**, 2673 (2004); J. Chem. Phys. **121**, 4068 (2004); J. Phys. Chem. A **108**, 8495 (2004); Y. Zhang, A. Wu, X. Xu, and Y. J. Yan, Chem. Phys. Lett. **421**, 383 (2006).
[21] Y. Zhao, N. E. Schultz, and D. G. Truhlar, J. Chem. Phys. **123**, 161103 (2005); Y. Zhao and D. G. Truhlar, J. Phys. Chem. A **109**, 5656 (2005); Y. Zhao, N. E. Schultz, and D. G. Truhlar, J. Chem. Theory Comput. **2**, 364 (2006); Y. Zhao and D. G. Truhlar, J. Chem. Phys. **125**, 194101 (2006); Y. Zhao, N. González-García, and D. G. Truhlar, J. Phys. Chem. A **109**, 2012 (2005).
[22] Y. Zhao and D. G. Truhlar, Theor. Chem. Acc. **120**, 215 (2008).
[23] Y. Zhao, B. J. Lynch, and D. G. Truhlar, J. Phys. Chem. A **108**, 4786 (2004).
[24] S. Grimme, J. Chem. Phys. **124**, 034108 (2006); T. Schwabe and S. Grimme, Phys. Chem. Chem. Phys. **9**, 3397 (2007).
[25] A. Tarnopolsky, A. Karton, R. Sertchook, D. Vuzman, and J. M. L. Martin, J. Phys. Chem. A **112**, 3 (2008); A. Karton, A. Tarnopolsky, J.-F. Lamère, G. C. Schatz, and J. M. L. Martin, *ibid.* **112**, 12868 (2008).
[26] J. C. Sancho-García and A. P. Pérez-Jiménez, J. Chem. Phys. **131**, 084108 (2009).
[27] D. C. Graham, A. S. Menon, L. Goerigk, S. Grimme, and L. Radom, J. Phys. Chem. A **113**, 9861 (2009).
[28] Y. Zhang, X. Xu, and W. A. Goddard III, Proc. Natl. Acad. Sci. U.S.A. **106**, 4963 (2009); I. Y. Zhang, J. M. Wu, Y. Luo, and X. Xu, J. Chem. Theory Comput. **6**, 1462 (2010); I. Y. Zhang, J. M. Wu, and X. Xu, Chem. Commun. (Cambridge) **2010**, 3057.
[29] J. P. Perdew, A. Ruzsinszky, J. M. Tao, V. N. Staroverov, G. E. Scuseria, and G. I. Csonka, J. Chem. Phys. **123**, 062201 (2005); A. Ruzsinszky, J. P. Perdew, and G. I. Csonka, J. Chem. Theory Comput. **6**, 127 (2010); L. A. Constantin, J. M. Pitarke, J. F. Dobson, A. Garcia-Lekue, and J. P. Perdew, Phys. Rev. Lett. **100**, 036401 (2008); J. F. Dobson, J. Wang, and T. Gould, Phys. Rev. B **66**, 081108 (2002).
[30] W. Kohn, A. D. Becke, and R. G. Parr, J. Phys. Chem. **100**, 12974 (1996).
[31] D. C. Langreth and J. P. Perdew, Phys. Rev. B **15**, 2884 (1977); O. Gunnarsson and B. I. Lundqvist, *ibid.* **13**, 4274 (1976); J. P. Perdew, M. Ernzerhof, and K. Burke, J. Chem. Phys. **105**, 9982 (1996).
[32] P. J. Stephens, F. J. Devlin, C. F. Chabalowski, and M. J. Frisch, J. Phys. Chem. **98**, 11623 (1994).
[33] J. M. Wu and X. Xu, J. Chem. Phys. **127**, 214105 (2007); J. Comput. Chem. **30**, 1424 (2009); J. Chem. Phys. **129**, 164103 (2008).
[34] P. M. W. Gill, B. G. Johnson, J. A. Pople, and M. J. Frisch, Chem. Phys. Lett. **197**, 499 (1992).
[35] N. X. Wang and A. K. Wilson, J. Phys. Chem. A **107**, 6720 (2003); Mol. Phys. **103**, 345 (2005); B. P. Prascher and A. K. Wilson, *ibid.* **105**, 2899 (2007).
[36] F. Jensen, J. Chem. Phys. **115**, 9113 (2001); **116**, 3502 (2002); **116**, 7372 (2002).
[37] A. C. Scheiner, J. Baker, and J. W. Andzelm, J. Comput. Chem. **18**, 775 (1997).
[38] A. D. Boese, J. M. L. Martin, and N. C. Handy, J. Chem. Phys. **119**, 3005 (2003).
[39] X. L. Wang and X. Xu, J. Phys. Chem. C **113**, 7843 (2009); N. X. Lu, G. Fu, X. Xu, and H. L. Wan, J. Chem. Phys. **128**, 034702 (2008); X. L. Wang and X. Xu, J. Phys. Chem. C **111**, 16974 (2007); G. Fu, X. Xu, and H. L. Wan, Catal. Today **117**, 133 (2006); G. Fu, X. Xu, X. Lu, and H. L. Wan, J. Phys. Chem. B **109**, 6416 (2005); X. Xu and W. A. Goddard III, Proc. Natl. Acad. Sci. U.S.A. **99**, 15038 (2002); X. Xu, S. Y. Kang, and T. Yamabe, Phys. Rev. Lett. **88**, 076106 (2002); X. Lu, X. Xu, N. Q. Wang, and Q. E. Zhang, J. Phys. Chem. B **103**, 3373 (1999).
[40] K. Wolinski and P. Pulay, J. Chem. Phys. **118**, 9497 (2003).
[41] L. A. Curtiss, K. Raghavachari, P. C. Redfern, and J. A. Pople, J. Chem. Phys. **112**, 7374 (2000); **109**, 7764 (1998); L. A. Curtiss, K. Raghavachari, G. W. Trucks, and J. A. Pople, *ibid.* **94**, 7221 (1991).
[42] M. J. Frisch, J. A. Pople, and J. S. Binkley, J. Chem. Phys. **80**, 3265 (1984); R. Krishnan, J. S. Binkley, R. Seeger, and J. A. Pople, *ibid.* **72**, 650 (1980).
[43] M. R. Nyden and G. A. Petersson, J. Chem. Phys. **75**, 1843 (1981); G. A. Petersson and M. A. Al-Laham, *ibid.* **94**, 6081 (1991); G. A. Petersson, T. G. Tensfeldt, and J. A. Montgomery, Jr., *ibid.* **94**, 6091 (1991); J. A. Montgomery, Jr., J. W. Ochterski, and G. A. Petersson, *ibid.* **101**, 5900 (1994).
[44] A. Görling and M. Levy, Phys. Rev. B **47**, 13105 (1993).
[45] M. J. Frisch, G. W. Trucks, H. B. Schlegel *et al.*, GAUSSIAN 03, Revision D.01, Gaussian, Inc., Wallingford CT, 2004.
[46] T. Helgaker, T. A. Ruden, P. Jørgensen, J. Olsen, and W. Klopper, J. Phys. Org. Chem. **17**, 913 (2004); D. Feller and E. R. Davidson, in *Reviews in Computational Chemistry*, edited by K. B. Lipkowitz and B. D. Boyd (VCH, New York, 1990), Vol. 1; T. Helgaker and P. R. Taylor, in *Modern Electronic Structure Theory, Part II*, edited by D. R. Yarkony (World Scientific, Singapore, 1995), p. 725; T. H. Dunning, Jr., K. A. Peterson, and D. E. Woon, *Encyclopedia of Computational Chemistry*, edited by P. v. R. Schleyer, N. L. Allinger, T. Clark, J. Gasteiger, P. A. Kollman, H. F. Schaefer III, and P. R. Schreiner (Wiley, Chichester, 1998), Vol. 1, p. 88.
[47] I. Y. Zhang, Y. Luo, and X. Xu, J. Chem. Phys. **132**, 194105 (2010).
[48] T. H. Dunning, Jr., J. Chem. Phys. **90**, 1007 (1989); R. A. Kendall, T. H. Dunning, Jr., and R. J. Harrison, *ibid.* **96**, 6796 (1992); D. E. Woon and T. H. Dunning, Jr., *ibid.* **99**, 3730 (1993).
[49] See supplementary material at http://dx.doi.org/10.1063/1.3488649 for Tables S1–S45.
[50] F. Jensen, Theor. Chem. Acc. **113**, 267 (2005); A. Karton and J. M. L. Martin, *ibid.* **115**, 330 (2006); S. Shahbazian and M. Zahedi, *ibid.* **113**, 152 (2005).
[51] A. Köhn and D. P. Tew, J. Chem. Phys. **132**, 024101 (2010).
[52] J. G. Hill, K. A. Peterson, G. Knizia, and H. Werner, J. Chem. Phys. **131**, 194105 (2009).
[53] D. Yamaki, H. Koch, and S. Ten-no, J. Chem. Phys. **127**, 144104 (2007).
[54] W. Klopper, B. Ruscic, D. P. Tew, F. A. Bischoff, and S. Wolfsegger, Chem. Phys. **356**, 14 (2009).
[55] C. E. Check and T. M. Gilbert, J. Org. Chem. **70**, 9828 (2005); E. I. Izgorodina, M. L. Coote, and L. Radom, J. Phys. Chem. A **109**, 7558 (2005).
[56] X. Wu, M. C. Vargas, S. Nayak, V. Lotrich, and G. Scoles, J. Chem. Phys. **115**, 8748 (2001); Q. Wu and W. Yang, *ibid.* **116**, 515 (2002); S. Grimme, J. Comput. Chem. **25**, 1463 (2004); P. Jurečka, J. Cerny, P. Hobza, and D. R. Salahub, *ibid.* **28**, 555 (2007).
[57] A. Halkier, H. Koch, P. Jørgensen, O. Christiansen, I. M. B. Nielsen, and T. Helgaker, Theor. Chem. Acc. **97**, 150 (1997); H. Valdés, V. Klusak, M. Pitonak, O. Exner, I. Stary, P. Hobza, and L. Rulisek, J. Comput. Chem. **29**, 861 (2008); P. Jurečka, J. Šponer, J. Černy, and P. Hobza, Phys. Chem. Chem. Phys. **8**, 1985 (2006).
[58] R. A. Distasio, Jr., and M. Head-Gordon, Mol. Phys. **105**, 1073 (2007).
[59] M. O. Sinnokrot, E. F. Valeev, and C. D. Sherrill, J. Am. Chem. Soc. **124**, 10887 (2002); C. D. Sherrill, T. Takatani, and E. G. Hohenstein, J. Phys. Chem. A **113**, 10146 (2009); S. Tsuzuki and H. P. Lüthi, J. Chem. Phys. **114**, 3949 (2001); S. K. Min, E. C. Lee, H. M. Lee, D. Y. Kim, D. Kim, and K. S. Kim, J. Comput. Chem. **29**, 1208 (2008); S. B. Huh and J. S. Lee, J. Chem. Phys. **118**, 3035 (2003).
[60] A. D. Boese, J. M. L. Martin, and W. Klopper, J. Phys. Chem. A **111**, 11122 (2007).

[61] Á. Vázquez-Mayagoitia, C. D. Sherrill, E. Aprà, and B. G. Sumpter, J. Chem. Theory Comput. **6**, 727 (2010).
[62] A. Schäfer, H. Horn, and R. Ahlrichs, J. Chem. Phys. **97**, 2571 (1992).
[63] P. C. Hariharan and J. A. Pople, Theor. Chim. Acta **28**, 213 (1973).
[64] N. Godbout, D. R. Salahub, J. Andzelm, and E. Wimmer, Can. J. Chem. **70**, 560 (1992).
[65] R. D. Adamson, P. M. W. Gill, and J. A. Pople, Chem. Phys. Lett. **284**, 6 (1998).
[66] B. J. Lynch, P. L. Fast, M. Harris, and D. G. Truhlar, J. Phys. Chem. A **104**, 4811 (2000).